Enhancing the Japanese WordNet

Francis Bond,[†] **Hitoshi Isahara,**[‡] **Sanae Fujita,**[♡]
Kiyotaka Uchimoto,[†] **Takayuki Kuribayashi**[†] and **Kyoko Kanzaki**[‡]
[†] NICT Language Infrastructure Group, [‡] NICT Language Translation Group
<bond@ieee.org,{isahara,uchimoto,kuribayashi,kanzaki}@nict.go.jp>
[♡] Sanae Fujita, NTT Communications Science Laboratory
<sanae@kecl.cslab.ntt.co.jp>

Abstract

The Japanese WordNet currently has 51,000 synsets with Japanese entries. In this paper, we discuss three methods of extending it: increasing the cover, linking it to examples in corpora and linking it to other resources (SUMO and GoiTaikei). In addition, we outline our plans to make it more useful by adding Japanese definition sentences to each synset. Finally, we discuss how releasing the corpus under an open license has led to the construction of interfaces in a variety of programming languages.

1 Introduction

Our goal is to make a semantic lexicon of Japanese that is both **accesible** and **usable**. To this end we are constructing and releasing the Japanese WordNet (`WN-Ja`) (Bond et al., 2008a).

We have almost completed the first stage, where we automatically translated the English and Euro WordNets, and are hand correcting it. We introduce this in Section 2. Currently, we are extending it in three main areas: the first is to add more concepts to the Japanese WordNet, either by adding Japanese to existing English synsets or by creating new synsets (§ 3). The second is to link the synsets to text examples (§ 4). Finally, we are linking it to other resources: the Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001), the Japanese semantic lexicon GoiTaikei (Ikehara et al., 1997), and a collection of illustrations taken from the Open ClipArt Library (Phillips, 2005) (§ 5).

2 Current State

Currently, the `WN-Ja` consists of 157,000 senses (word-synset pairs) 51,000 concepts (synsets) and 81,000 unique Japanese words (version 0.91). The relational structure (hypernym, meronym, domain, ...) is based entirely on the English WordNet 3.0 (Fellbaum, 1998). We have Japanese words for 43.0% of the synsets in the English WordNet. Of these synsets, 45% have been checked by hand, 8% were automatically created by linking through multiple languages and 46% were automatically created by adding non-ambiguous translations, as described in Bond et al. (2008a). There are some 51,000 synsets with Japanese candidate words that have not yet been checked. For up-to-date information on `WN-Ja` see: `nlpwww.nict.go.jp/wn-ja`.

An example of the entry for the synset 02076196-n is shown in Figure 1. Most fields come from the English WordNet. We have added the underlined fields (Ja Synonyms, Illustration, links to GoiTaikei, SUMO) and are currently adding the translated definition (Def (Ja)). In the initial automatic construction there were 27 Japanese words associated with the synset,[1] including many inappropriate translations for other senses of *seal* (e.g., 判こ *hanko* "stamp"). These were reduced to three after checking: アザラシ, 海豹 *azarashi* "seal" and シール *shi-ru* "seal". Synsets with ? in their names are those for which there is currently no Japanese entry in the WordNet.

The main focus of this year's work has been this manual trimming of badly translated words. The result is a WordNet with a reasonable coverage of common Japanese words. The precision per sense is just over 90%. We have aimed at high coverage at the cost of precision for two reasons: (i) we think that the WordNet must have a rea-

[1] アザラシ, シール, スタンプ, 上封, 判, 判こ, 判子, 刻印, 加判, 印, 印判, 印形, 印章, 印鎰, 印鑑, 印鑰, 印顆, 墨引, 墨引き, 封, 封じ目, 封印, 封目, 封着, 封緘, 押し 手, 押印, 押手, 押捺, 捺印, 極印, 海豹, 版行, 符節, 緘, 証印, 調印

Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009, pages 1–8,
Suntec, Singapore, 6-7 August 2009.

sonable coverage to be useful for NLP tasks and (ii) we expect to continue refining the accuracy over the following years. Our strategy is thus different from Euro WordNet (Vossen, 1998), where initial emphasis was on building a consistent and complete upper ontology.

3 Increasing Coverage

We are increasing the coverage in two ways. The first is to continue to manually correct the automatically translated synsets. This is being done both by hand, as time permits, and also by comparing against other resources such as GoiTaikei and Wikipedia. When we check for poor candidates, we also add in missing words as they occur to us.

More interestingly, we wish to add synsets for Japanese concepts that may not be expressed in the English WordNet. To decide which new concepts to add, we will be guided by the other tasks we are doing: annotation and linking. We intend to create new synsets for words found in the corpora we annotate that are not currently covered, as well as for concepts that we want to link to. An example for the first is the concept 御飯 *gohan* "cooked rice", as opposed to the grain 米 *kome* "rice". An example of the second is シングル *shinguru* "single: a song usually extracted from a current or upcoming album to promote the album". This is a very common hypernym in Wikipedia but missing from the English WordNet.

As far as possible, we want to coordinate the creation of new synsets with other projects: for example KorLex: the Korean WordNet already makes the cooked rice/grain distinction, and the English WordNet should also have a synset for this sense of *single*.

4 Text Annotation

We are in the process of annotating four texts (Table 1). The first two are translations of WordNet annotated English Texts (SemCor and the WordNet definitions), the third is the Japanese newspaper text that forms the Kyoto Corpus and the fourth is an open corpus of bilingual Japanese-English sentences (Tanaka). In 2009, we expect to finish translating and annotate all of SemCor, translate the WordNet definitions and start annotation on the Kyoto and Tanaka Corpora.

Name	Sentences	Words	Content Words
SemCor	12,842	224,260	120,000
Definitions	165,977	1,468,347	459,000
Kyoto	38,383	969,558	527,000
Tanaka	147,190	1,151,892	360,000

Table 1: Corpora to be Sense Tagged

This annotation is essential for finding missing senses in the Japanese WordNet, as well as getting the sense distributions that are needed for supervised word sense disambiguation.

4.1 SemCor

SemCor is a textual corpus in which words have been both syntactically and semantically tagged. The texts included in SemCor were extracted from the Brown corpus (Francis and Kucera, 1979) and then linked to senses in the English WordNet. The frequencies in this corpus were used to give the sense frequencies in WordNet (Fellbaum, 1998). A subset of this corpus (MultiSemCor) was translated into Italian and used as a corpus for the Italian WordNet (Bentivogli et al., 2004). We are translating this subset into Japanese.

In the same way as Bentivogli et al. (2004), we are exploiting Cross-Language Annotation Transfer to seed the Japanese annotation. For example, consider (1)[2]. The content words *answer, was, simple, honest* are tagged in SemCor. They can be aligned with their translations 答え *kotae* "answer", 簡単 *kantan* "simple", 率直 *socchoku* "honest" and だった *datta* "was". This allows us to tag the Japanese translation with the same synsets as the English, and thus disambiguate them.

(1) His answer$_i$ was$_j$ simple$_k$ but honest$_l$.
答え$_i$ は 簡単$_k$ ながらも 率直$_l$ な ものだった$_j$ 。

However, just because all the English words have sysnets in WordNet, it is not always the case for the translations. For example, the English phrase *last night* can be translated into 前夜 *zen'ya* "last-night". Here the two English words (and synsets) link to a single Japanese

[2]Sentence 96 in b13.

Synset	02076196-n
Synonyms	ja 海豹, アザラシ, シール en seal$_9$ fr phoque
Illustration	 animal/seal.png
Def (en)	"any of numerous marine mammals that come on shore to breed; chiefly of cold regions"
Def (ja)	「繁殖のために岸に上がる海洋性哺乳動物の各種；主に寒帯地域に」
Hypernyms	アシカ亜目/pinniped
Hyponyms	？/crabeater_seal ？/eared_seal 海驢/earless_seal
GoiTaikei	⟨⟨537:beast⟩⟩
SUMO	⊂ Carnivore

Figure 1: Example Entry for Seal/海豹

word which has no suitable synset in the English WordNet. In this case, we need to create a new synset unique to the Japanese WordNet.[3]

We chose a translated SemCor as the basis of annotation for two main reasons: (i) the corpus can be freely redistributed — we expect the glosses to be useful as an aligned corpus of Japanese-English-Italian and (ii) it has other annotations associated with it: Brown corpus POS annotation, Penn Treebank syntactic annotation.

4.2 WordNet Definitions

Our second translated corpus is formed from the WordNet definitions (and example sentences) themselves (e.g., the **def** field shown in Figure 1). The English definitions have been annotated with word senses in the *Princeton WordNet Gloss Corpus*. In the same way that we do for SemCor, we are translating the definitions and examples, and using the existing annotation to seed our annotation.

Using the definitions as the base for a sense annotated corpus is attractive for the following reasons: (i) the translated corpus can be freely redistributed — we expect the definitions to be useful as an aligned corpus and also to be useful for many other open lexicons; (ii) the definitions are useful for Japanese native speakers using the WordNet, (iii) the definitions are useful for unsupervised sense disambiguation techniques such as LESK (Baldwin et al., 2008); (iv) other projects have also translated synset definitions (e.g. Spanish and Korean), so we can hope to create a multilingual corpus here as well and (v) the definitions can be used as a machine readable dictionary, and various information extracted from there (Barnbrook, 2002; Nichols et al., 2006)

4.3 Kyoto Text Corpus

The Kyoto Text Corpus consists of newspaper text from the Mainichi Newspaper (1995), segmented and annotated with Japanese POS tags and dependency trees (Kurohashi and Nagao, 2003). The corpus is made up of two parts. The first consists of 17 full days of articles and the second of one year's editorials. We hope to annotate at least parts of it during 2009.

Even though the Kyoto Text Corpus is not freely redistributable, we have chosen to annotate it due to the wealth of annotation associated with it: dependency trees, predicate-argument relations and co-reference (Iida et al., 2007), translations into English and Chinese (Uchimoto et al., 2004) and sense annotations from the Hinoki project (Bond et al., 2006). We also felt it was important to tag some native Japanese text, not only translated text.

4.4 Tanaka Corpus

Finally, we will also tag the Tanaka Corpus, an open corpus of Japanese-English sentence pairs compiled by Professor Yasuhito Tanaka at Hyogo University and his students (Tanaka, 2001) and released into the public domain. The corrected version we use has around 140,000 sentence pairs.

This corpus is attractive for several reasons. (i) it is freely redistributable; (ii) it has been indexed to entries in the Japanese-English dictio-

[3] Arguably, the fact that one says *last night* (not *yesterday night*) for the night proceeding today and *tomorrow night* (not *next night*) for the night following today suggests that these multi-word expressions are lexicalized and synsets should be created for them in the English WordNet. However, in general we expect to create some synsets that will be unique to the Japanese WordNet.

nary JMDict (Breen, 2003); (iii) part of it has also been used in an open HPSG-based treebank (Bond et al., 2008b); (iv) further, translations in other languages, most notably French, have been added by the TATOEBA project.[4] Our plan is to tag this automatically using the tools developed for the Kyoto corpus annotation, and then to open the data to the community for refinement. We give a typical example sentence in (2).

(2) あの木の枝に数羽の鳥がとまっている。
"Some birds are sitting on the branch of that tree." (en)
"Des oiseaux se reposent sur la branche de cet arbre." (fr)

5 Linking to other resources

We currently link the Japanese WordNet to three other resources: the Suggested Upper Merged Ontology; GoiTaikei, a Japanese Lexicon; and a collection of pictures from the Open Clip Art Library (OCAL: Phillips (2005)).

For SUMO we used existing mappings. For the other resources, we find confident matches automatically and then generalize from them. We find matches in three ways:

MM Monosemous monolingual matches
e.g. *cricket bat* or 海豹 *azarashi* "seal"

MB Monosemous bilingual matches
e.g. ⟨海豹↔*seal*⟩

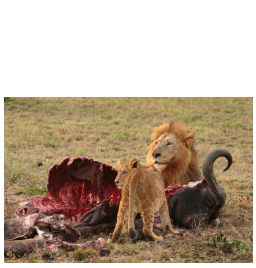
HH Hypernym/Hyponym pairs
e.g. ⟨seal ⊂ mammal⟩

We intend to use the same techniques to link other resources, such as the concepts from the EDR lexicon (EDR, 1990) and the automatically extracted hypernym-hyponym links from Torishiki-kai (Kuroda et al., 2009).

5.1 SUMO

The Suggested Upper Merged Ontology (SUMO) is a large formal public ontology freely released by the IEEE (Niles and Pease, 2001).

Because the structure of the Japanese WordNet is closely linked to that of the English WordNet, we were able to take advantage of the existing mappings from the English WordNet to SUMO. There are 102,669 mappings from SUMO to WordNet: 3,593 equivalent, 10,712 where the WordNet synset subsumes the SUMO concept, 88,065 where the SUMO concept subsumes the WordNet concept, 293 where the negation of the SUMO concept subsumes the WordNet synset and 6 where the negation of the SUMO concept is equivalent to the WordNet synset. According to the mapping, synset 02076196-n 海豹 *azarashi* "seal", shown in Figure 1 is subsumed by the SUMO concept ⟨⟨Carnivore⟩⟩. There is no link between *seal* and *carnivore* in WordNet, which shows how different ontologies can complement each other.


Carnivore


Business Competition

Figure 2: SUMO illustrations

Linking to SUMO also allowed us to use the SUMO illustrations.[5] These consist of 12,237 links linking 4,607 concepts to the urls of 10,993 illustrations. These are mainly taken from from Wikimedia (`upload.wikimedia.org`), with around 1,000 from other sources. The pictures can be linked quite loosely to the concepts. For example, ⟨⟨Carnivore⟩⟩ is illustrated by a lion eating meat, and ⟨⟨BusinessCompetition⟩⟩ by a picture of Wall Street.

As we wanted our illustrations to be more concrete, we only use SUMO illustrations where the SUMO-WordNet mapping is equivalence. This gave 4,384 illustrations for 999 synsets.

5.2 GoiTaikei

Linking Goi-Taikei, we used not only the Japanese dictionary published in Ikehara et al. (1997), but also the Japanese-English dictionary used in the machine translation system ALT-J/E (Ikehara et al., 1991). We attempted to match synsets to semantic categories by matching the

[4] `wwwcyg.utc.fr/tatoeba/`

[5] Available at `http://sigmakee.cvs.sourceforge.net/viewvc/sigmakee/KBs/pictureList.kif`, thanks to Adam Pease for letting us know about them.

Japanese, English and English-Japanese pairs to unambiguous entries in Goi-Taikei. For example, the synset shown in Figure 1 was automatically assigned the semantic category ⟨⟨537:beast⟩⟩, as 海豹 appears only once in WN-Ja, with the synset shown, and once in the Japanese dictionary for ALT-J/E with a single semantic category.

We are currently evaluating our results against an earlier attempt to link WordNet and GoiTaikei that also matched synset entries to words in Goi-Taikei (Asanoma, 2001), but did not add an extra constraint (that they must be either monosemous or match as a hypernym-hyponym pair).

Once we have completed the mapping, we will use it to check for inconsistencies in the two resources.

5.3 Open ClipArt Library

In order to make the sense distinctions more visible we have semi-automatically linked synsets to illustrations from the Open Clip Art Library (OCAL: Phillips (2005)) using the mappings produced by Bond et al. (2008a).

We manually checked the mappings and added a goodness score. Illustrations are marked as:

3 the best out of multiple illustrations

2 a good illustration for the synset

1 a suitable illustration, but not perfect
This tag was used for black and white images, outlines, and so forth.

After the scoring, there were 874 links for 541 synsets (170 scored 1, 642 scored 2 and 62 scored 3). This is only a small subset of illustrations in OCAL and an even smaller proportion of wordnet. However, because any illustrated synset also (in theory) illustrates its hypernyms, we have indirectly illustrated far more than 541 synsets: these figures are better than they seem.

There are far fewer OCAL illustrations than the SUMO linked illustrations. However, they are in general more representative illustrations (especially those scored 2 and above), and the source of the clipart is available as SVG source so it is easy to manipulate them. We think that this makes them particularly useful for a variety of tasks. One is pedagogical — it is useful to have pictures in learners' dictionaries. Another is in cross-cultural communication - for example in Pangea, where children use pictons (small concept representing pictures) to write messages (Takasaki and Mori, 2007).

The OCAL illustrations mapped through WordNet to 541 SUMO concepts. We have given these links to the SUMO researchers.

6 Interfaces

We released the Japanese WordNet in three formats: tab-delimited text, XML and as an SQLite database. The license was the same as English WordNet. This is a permissive license, the data can be reused within proprietary software on the condition that the license is distributed with that software (similar to the MIT X license). The license is also GPL-compatible, meaning that the GPL permits combination and redistribution with software that uses it.

The tab delimited format consists of just a list of synsets, Japanese words and the type of link (hand, multi-lingual or monosemous):

02076196-n	海豹	hand
02076196-n	アザラシ	hand
02076196-n	シール	hand

We also output in WordNet-LMF (Francopoulo et al., 2006; Soria et al., 2009), to make the program easily available for other WordNet researchers. In this case the synset structure was taken from the English WordNet and the lemmas from the Japanese WordNet. Because of the incomplete coverage, not all synsets contain lemmas. This format is used by the Kyoto Project, and we expect it to become the standard exchange format for WordNets (Vossen et al., 2008).

Finally, we also created an SQL database. This contains information from the English WordNet, the Japanese WordNet, and links to illustrations. We chose SQLite,[6] a self-contained, zero-configuration, SQL database engine whose source code is in the public domain. The core structure is very simple with six tables, as shown in Figure 3.

As we prepared the release we wrote a perl module for a basic interface. This was used to develop a web interface: Figure 4 shows a screenshot.

[6] http://www.sqlite.org

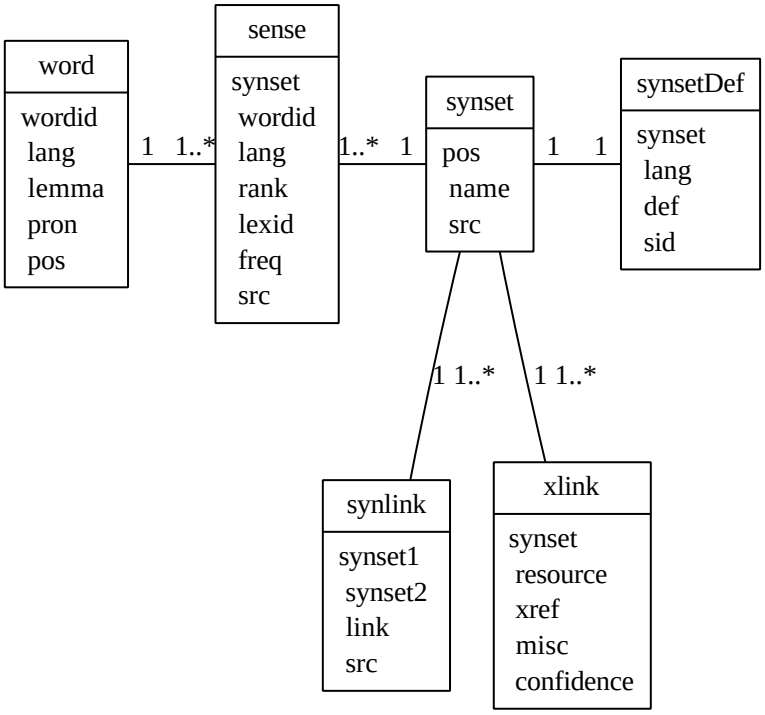


Figure 3: Database Schema

Figure 4: Web Search Screenshot

7 Discussion

In contrast to earlier WordNets, the Japanese WordNet was released with two known major imperfections: (i) the concept hierarchy was entirely based on English with no adaptation to Japanese and (ii) the data was released with some unchecked automatically created entries. The result was a WordNet that did not fully model the lexical structure of Japanese and was known to contain an estimated 5% errors. The motivation behind this was twofold. Firstly, we wanted to try and take advantage of the open source model. If the first release was good enough to be useful, we hoped to (a) let people use it and (b) get feedback from them which could then be incorporated into the next release. This is the strategy known as *release early, release often* (Raymond, 1999).

Secondly, we anticipated the most common use of the WordNet to be in checking whether one word is a hypernym of another. In this case, even if one word is wrong, it is unlikely that the other will be, so a small percentage of errors should be acceptable.

From the practical point of view, the early release appears to have been a success. The SQL database proved very popular, and within two weeks of the first release someone produced a python API. This was soon followed by interfaces in java, ruby, objective C and gauche. We also received feedback on effective indexing of the database and some corrections of entries — these have been included in the most recent release (0.91).

The data from the Japanese WordNet has already been incorporated into other projects. The first was the Multi-Lingual Semantic Network (MLSN) (Cook, 2008) a WordNet based network of Arabic, Chinese, English, German and Japanese. Because both the Japanese WordNet and MLSN use very open licenses, it is possible to share entries directly. We have already received useful feedback and over a thousand new entries from MLSN. The second project using our data is the Asian WordNet (Charoenporn et al., 2008). They have a well developed interface for collaborative development of linguistic resources, and we hope to get corrections and additions from them in the future. Another project using the Japanese WordNet data is the Language Grid (Ishida, 2006) which offers the English and Japanese WordNets as concept dictionaries.

We have also been linked to from other resources. The Japanese-English lexicon project JMDict (Breen, 2004) now links to the Japanese WordNet, and members of that project are using WordNet to suggest new entries. We used JMDict in the first automatic construction stage, so it is particularly gratifying to be able to help JMDict in turn.

Finally, we believe that data about language should be shared — language is part of the common heritage of its speakers. In our case, the Japanese WordNet was constructed based on the work that others made available to us and thus we had a moral obligation to make our results freely available to others. Further, projects that create WordNets but do not release them freely hinder research on lexical semantics in that language — people cannot use the unreleased resource, but it is hard to get funding to duplicate something that already exists.

In future work, in addition to the planned extensions listed here, we would like to work on the following: Explicitly marking lexical variants; linking to instances in Wikipedia; adding derivational and antonym links; using the WordNet for word sense disambiguation.

8 Conclusion

This paper presents the current state of the Japanese WordNet (157,000 senses, 51,000 concepts and 81,000 unique Japanese words, with links to SUMO, Goi-Taikei and OCAL) and outlined our plans for further work (more words, links to corpora and other resources). We hope that `WN-Ja` will become a useful resource not only for natural language processing, but also for language education/learning and linguistic research.

References

Naoki Asanoma. 2001. Alignment of ontologies:wordnet and goi-taikei. In *NAACL Wokshop on WordNet & Other Lexical Resources*, pages 89–94. Pittsburgh, USA.

Timothy Baldwin, Su Nam Kim, Francis Bond, Sanae Fujita, David Martinez, and Takaaki Tanaka. 2008. MRD-based word sense disambiguation: Further extending Lesk. In *Proc. of the 3rd International Joint Conference on Natural Language Processing (IJCNLP-08)*, pages 775–780. Hyderabad, India.

Geoff Barnbrook. 2002. *Defining Language — A local*

grammar of definition sentences. Studies in Corpus Linguistics. John Benjamins.

Luisa Bentivogli, Pamela Forner, and Emanuele Pianta. 2004. Evaluating cross-language annotation transfer in the MultiSemCor corpus. In *20th International Conference on Computational Linguistics: COLING-2004*, pages 364–370. Geneva.

Francis Bond, Sanae Fujita, and Takaaki Tanaka. 2006. The Hinoki syntactic and semantic treebank of Japanese. *Language Resources and Evaluation*, 40(3–4):253–261. (Special issue on Asian language technology).

Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008a. Boot-strapping a WordNet using multiple existing WordNets. In *Sixth International conference on Language Resources and Evaluation (LREC 2008)*. Marrakech.

Francis Bond, Takayuki Kuribayashi, and Chikara Hashimoto. 2008b. Construction of a free Japanese treebank based on HPSG. In *14th Annual Meeting of the Association for Natural Language Processing*, pages 241–244. Tokyo. (in Japanese).

James W. Breen. 2003. Word usage examples in an electronic dictionary. In *Papillon (Multi-lingual Dictionary) Project Workshop*. Sapporo.

James W. Breen. 2004. JMDict: a Japanese-multilingual dictionary. In *Coling 2004 Workshop on Multilingual Linguistic Resources*, pages 71–78. Geneva.

Thatsanee Charoenporn, Virach Sornlerlamvanich, Chumpol Mokarat, and Hitoshi Isahara. 2008. Semi-automatic compilation of Asian WordNet. In *14th Annual Meeting of the Association for Natural Language Processing*, pages 1041–1044. Tokyo.

Darren Cook. 2008. MLSN: A multi-lingual semantic network. In *14th Annual Meeting of the Association for Natural Language Processing*, pages 1136–1139. Tokyo.

EDR. 1990. Concept dictionary. Technical report, Japan Electronic Dictionary Research Institute, Ltd.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

W. Nelson Francis and Henry Kucera. 1979. *BROWN CORPUS MANUAL*. Brown University, Rhode Island, third edition.

Gil Francopoulo, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet, and Claudia Soria. 2006. Lexical markup framework (LMF). In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy.

Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *ACL Workshop: Linguistic Annotation Workshop*, pages 132–139. Prague.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. *Goi-Taikei — A Japanese Lexicon*. Iwanami Shoten, Tokyo. 5 volumes/CDROM.

Satoru Ikehara, Satoshi Shirai, Akio Yokoo, and Hiromi Nakaiwa. 1991. Toward an MT system without pre-editing — effects of new methods in **ALT-J/E** —. In *Third Machine Translation Summit: MT Summit III*, pages 101–106. Washington DC.

Toru Ishida. 2006. Language grid: An infrastructure for intercultural collaboration. In *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pages 96–100. (keynote address).

Kow Kuroda, Jae-Ho Lee, Hajime Nozawa, Masaki Murata, and Kentaro Torisawa. 2009. Manual cleaning of hypernyms in Torishiki-Kai. In *15th Annual Meeting of The Association for Natural Language Processing*, pages C1–3. Tottori. (in Japanese).

Sadao Kurohashi and Makoto Nagao. 2003. Building a Japanese parsed corpus — while improving the parsing system. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 14, pages 249–260. Kluwer Academic Publishers.

Eric Nichols, Francis Bond, Takaaki Tanaka, Sanae Fujita, and Daniel Flickinger. 2006. Robust ontology acquisition from multiple sources. In *Proceedings of the 2nd Workshop on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 10–17. Sydney.

Ian Niles and Adam Pease. 2001. Towards a standard upper ontology. In Chris Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*. Maine.

Jonathan Phillips. 2005. Introduction to the open clip art library. `http://rejon.org/media/writings/ocalintro/ocal_intro_phillips.html`. (accessed 2007-11-01).

Eric S. Raymond. 1999. *The Cathedral & the Bazaar*. O'Reilly.

Claudia Soria, Monica Monachini, and Piek Vossen. 2009. Wordnet-LMF: fleshing out a standardized format for wordnet interoperability. In *Second International Workshop on Intercultural Collaboration (IWIC-2009)*. Stanford.

Toshiyuki Takasaki and Yumiko Mori. 2007. Design and development of a pictogram communication system for children around the world. In *First International Workshop on Intercultural Collaboration (IWIC-2007)*, pages 144–157. Kyoto.

Yasuhito Tanaka. 2001. Compilation of a multilingual parallel corpus. In *Proceedings of PACLING 2001*, pages 265–268. Kyushu.

Kiyotaka Uchimoto, Yujie Zhang, Kiyoshi Sudo, Masaki Murata, Satoshi Sekine, and Hitoshi Isahara. 2004. Multilingual aligned parallel treebank corpus reflecting contextual information and its applications. In Gilles Sérasset, editor, *COLING 2004 Multilingual Linguistic Resources*, pages 57–64. COLING, Geneva, Switzerland.

P Vossen, E. Agirre, N. Calzolari, C. Fellbaum, S. Hsieh, C. Huang, H. Isahara, K. Kanzaki, A. Marchetti, M. Monachini, F. Neri, R. Raffaelli, G. Rigau, and M. Tescon. 2008. KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco.

Piek Vossen, editor. 1998. *Euro WordNet*. Kluwer.